DOCUMENT RESUME

ED 481 389 SP 041 735

AUTHOR	Walsh, Kate
TITLE	Teacher Certification Reconsidered: Stumbling for Quality. A Rejoinder.
INSTITUTION	Abell Foundation, Baltimore, MD.
PUB DATE	2001-11-00
NOTE	19p.; Written with contributing analysis by Michael Podgursky.
PUB TYPE	Opinion Papers (120) -- Reports - Descriptive (141)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Educational Research; Elementary Secondary Education; Higher Education; *Preservice Teacher Education; *Teacher Certification; Teacher Influence

ABSTRACT

This paper asserts that in a response to the Abell Foundation's "Teacher Certification Reconsidered, Stumbling for Quality," Linda Darling-Hammond mounts a considerable effort to discredit the report. It suggests that in doing so, she misrepresents the report's numerous facts and recommendations, shifting the debate off the primary concern of Abell's research: whether there is research which proves that certified teachers produce greater student achievement than do uncertified teachers. The paper also asserts that Darling-Hammond's response cites poor-quality studies. The paper specifically addresses each of Darling-Hammond's charges and assertions, providing a technical analysis that explores what Darling-Hammond referred to as poor quality research. It also looks at Darling-Hammond's analysis of studies, corrects her assertions of errors within the studies, and challenges several general points made by Darling-Hammond (e.g., misrepresentations of conversations with researchers, dismissal of studies simply because they were old, recommendation that states collect verbal ability scores only on prospective teachers who have gone to schools of education, and lack of distinction between teachers who are not certified at all and teachers who are alternatively certified). (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 481 389

Teacher Certification Reconsidered: Stumbling for Quality

A Rejoinder

Kate Walsh
Senior Policy Analyst
The Abell Foundation

With contributing analysis by
Michael Podgursky
Chair, Department of Economics
University of Missouri-Columbia

November 2001

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Anne Culman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

OVERVIEW

In a lengthy response to The Abell Foundation's *Teacher Certification Reconsidered, Stumbling for Quality* (www.abell.org), Linda Darling-Hammond mounts a considerable effort to discredit the report. In doing so, she misrepresents the report's numerous facts and recommendations. She shifts the debate off the primary concern of Abell's research: does research exist proving that certified teachers produce greater student achievement than do uncertified teachers? She dismisses Abell's review of *every* study that she and others have cited on this subject, insisting that there are large numbers of studies still unexamined. She then continues to misrepresent the design, methodology and findings of the few studies that do meet basic scientific standards. Darling-Hammond revisits a few of the authors with whom Walsh had conversed, but nothing in her retelling of these conversations alters Abell s written analyses.

Even if one were to overlook the inferior design and methodologies that characterize the 19 studies cited by Darling-Hammond in her response[1] (down from over 200 that Darling-Hammond has referred to in previous writings), and which she claims demonstrate the value of teacher certification, these studies have little to offer. Viewed through any lens, they certainly do not provide sufficient evidence to justify the current policy of 50 states that bar teachers from the classroom who are not certified. The issue at hand here is not whether schools of education offer some helpful and valuable coursework. They undoubtedly do. The issue is whether individuals who have not taken any education coursework (valuable or not) are at such a disadvantage that they should not be allowed to begin teaching. The evidence that would justify such a restriction is simply not there.

But the poor quality of the studies cited by Darling-Hammond cannot and should not be overlooked. Unfortunately, unlike other disciplines, the education field does not self-police, requiring little to no technical review prior to publication. Trying to explain to the lay reader *after* something is published why it should never have been published is a challenge and one that I believe Abell's study did with great competence. But for this rejoinder, we include a more technical analysis of the quality of these studies, supplied by Dr. Michael Podgursky. Dr. Podgursky is the Chair of the Economics Department at the University of Missouri, Columbia with strong credentials in the research on teacher quality. His expertise in this field should (but will not, of course) allay any questions as to the accuracy of Abell's analysis.

Darling-Hammond proves a formidable opponent in this debate, simply because her rules of engagement part from the norm. By way of illustration, read the following excerpt from her response:

---

[1] The 19 studies named by Darling-Hammond in her response are: National Reading Panel (2000); Wenglinsky (2000); Miller, McKenna and McKenna (1998); Goldhaber and Brewer (2000); Hawk, Coble and Swanson (1985); Fetler (1999); Begle (1972); Begle and Geeslin (1972) Monk (1994); Greenwald, Hedges and Laine (1996); Ferguson (1991); Strauss and Sawyer (1986); Ferguson and Womack (1993); Guyton and Farokhi (1987); Druva and Anderson (1983), Darling-Hammond (2000); Jelmberg (1995); Andrew and Schwab (1996); Denton and Peters (1998).

[Walsh] misreads the findings of [my 1999] study when she claims it found that, "certified teachers are only shown to have a significant effect on one out of six measurements." This is wrong: the proportion of well-qualified teachers (those with full certification and a major in the field) had a significantly positive effect in every one of the regression estimates.

As I read the above passage, I assumed that I had misread the table and was prepared to admit to an error. I went back to her 1999 study to check the facts, but still could not find an error in my interpretation. Again, I confirmed that Row 3 of the table reads "Percent of all teachers fully certified" followed by six sets of numbers. Only one of the six indicates that certification was significantly related to student achievement. Rereading her response, I then realized that she was now referring to a different measure, one not found in Row 3 but in Row 1 "Certified teachers who also had majors in the subject." On this measure, all six sets of numbers were in fact significant, but they were only significant because the certified teachers had majored in the subject. Without the major, only one of the six measures proved significant. My original statement was not wrong.

Darling-Hammond asks a reasonable question from her perspective. If we accept what Abell asserts about certification, how are teachers to learn what is known about how to teach well if there are no expectations, incentives or supports for them to do so? The answer though is clear: the same way new teachers learn now, but school districts will be a lot more deliberate about the need to mentor, train, and provide good staff development and be given more flexibility to decide who they are willing to train. Funds that flow into required coursework can be diverted to give new teachers a partial teaching load in their first year. The nation currently has a system in place that pretends that prospective teachers are adequately trained before they enter the classroom, with school districts then having to start the real training the day they enter the classroom. All Abell has done is to show that the regulatory policies established in 50 states are neither the most efficient nor the most effective means for ensuring high standards for entry into the teaching profession. In fact, they are often counterproductive to recruiting teachers of quality.

SPECIFIC CHARGES AND ASSERTIONS

On the following pages, each of Darling-Hammond's charges and assertions are specifically addressed. Dr. Podgursky provides the technical analysis that explores in more technical detail the poor quality of the research cited by Darling-Hammond.

DARLING-HAMMOND: *Maryland's **Resident Teacher Program** has been a "revolving door of under-prepared teachers recently targeted for discontinuation by the new superintendent of schools because of its high attrition rates and poor outcomes for children" (page 1)*

RESPONSE: This statement is inaccurate. The decision by the Baltimore City Public School CEO to stop targeted recruitment of the Resident Teacher program was not related to high attrition and poor outcomes of this program (neither of which has ever been even asserted to be the case). The reasons were pressure from the State and the desire to make budget cuts. Since the Abell report was released, the CEO has announced that she would be tripling the number of Resident Teachers in Baltimore City next year.

DARLING-HAMMOND: *Walsh did not include a large number of the studies cited as relevant to the question of teacher education effects.*

RESPONSE: In the course of her 69-page response, Darling-Hammond never names these studies. She adds only five studies to those she has previously cited in other works. Subsequently, I have looked at each of them. Two of them are immaterial because they do not discuss teacher certification and not one of the five alters any statement in the paper. It should be emphasized that I looked at every study that Darling-Hammond has cited advocating the case for teacher certification, so her assertion that there are a large number of studies that I did not consider is without merit. She asserts again on page 33 that Walsh dismisses "more fine grained studies from consideration" without naming such studies.

DARLING-HAMMOND: *The Abell report asserts that verbal ability and subject matter knowledge alone make a teacher effective (pages 5, 8).*

RESPONSE: This is an exaggeration of our statements regarding verbal ability and subject matter knowledge. The report states that there are only a few teacher attributes that have been shown to be **measurable** and that states are not currently measuring them. Because it is impossible to predict with any certainty who will be an effective teacher through the use of any assessment or by counting course titles (as is states' current practice), the report argues that the final, albeit subjective, judgment should be made by the person held accountable by state policies for raising student achievement: the school principal. There are clearly many attributes that are important for teachers to have, but most of them are not measurable and the report makes this point repeatedly. However, states could provide a tremendous service to school districts by providing the tools to measure verbal ability and content knowledge, which are important attributes a new teacher should possess.

DARLING-HAMMOND: *Walsh ignores evidence about importance of teachers' instruction for reading (page 5).*

RESPONSE: Darling-Hammond's refers to an excellent report by the National Reading Panel, though it contains no evidence that schools of education are making teachers more effective in reading instruction. It did not explore the relationship of pre-service teacher education with teacher's later effectiveness, due to the lack of rigorous studies on this point (National Reading Panel, page 13). More importantly, the need to train teachers in effective reading instruction is not an issue that advocates of teacher education can take on with any credibility, given their track record of disregarding the research. Take for example this excerpt from a textbook used by an NCATE-accredited school of education:

> "Who advocates the teaching of extensive and intensive phonics? Typically it is not reading researchers or educators, even those who advocate systematic phonics.... It is mainly laypersons --that is, those with no educational background in the process or the teaching of reading-- who advocate the extensive and intensive teaching of phonics. Typically, the impetus for teaching phonics extensively and intensively comes from certain leaders and their organizations among the political and religious Far Right.
>
> "What motivates such advocacy? Oddly enough, it may not necessarily be what proponents claim: namely, the desire to teach all children to read. A great deal of the force behind such advocacy seems to be the desire to promote a religious agenda and/or to maintain the socioeconomic status quo."[2]

This view underscores our argument that schools of education cannot hold a monopoly on teacher preparation.

Darling-Hammond: *Walsh misinterprets the composition of the TECAT (a teacher test used briefly in Texas; researcher Ronald Ferguson found a strong relationship between teacher scores on this test and their students' achievement). This test should rightfully be considered in part a test of professional knowledge. It provides evidence that knowledge about teaching and learning does in fact contribute to student achievement.*

RESPONSE: Darling-Hammond s portrayal of the TECAT is incorrect. The researcher, Ronald Ferguson calls the TECAT a test of basic literacy, not a test of professional knowledge. Even if it did actually test for professional knowledge, it does not speak well for the rigor of a formal teacher education program if 97% of the teachers who took this test passed it. The test did include 10 items which tested job related vocabulary. Teachers were asked on multiple choice questions if they could identify the right definition of such terms as standardized tests, classroom management, and certification. [3] Can anyone argue that knowledge of these terms requires 30 credit hours of education coursework or that knowledge of these words demonstrates a teacher s knowledge of teaching and learning ?

---

[2] Weaver, Constance (1994). *Reading Process and Practice: From Socio-Psycholinguistics to Whole Language.* Heinemann; Portsmouth, NH. From a chapter entitled "Phonics and Whole Language: From Politics to Research."

[3] Sample questions provided by the Texas State Board of Educator Certification.

DARLING-HAMMOND: *The Recent College Graduates Survey, which tracks college graduates into the labor market, found that the grade point averages of newly qualified teachers in 1990 were higher than those of the average college graduate.*

RESPONSE: GPAs are neither valid nor reliable measures by which to compare ability. Comparing the grade point averages of those in highly rigorous undergraduate programs such as pre-med with those in less rigorous programs is a misleading exercise. It has been shown that GPA's of education majors are on average substantially higher than other majors. The National Center for Education Statistics (NCES) offers useful findings on this point.[4] They note that education course grades only seem to vary between 3.0 and 4.0. The average grade nationwide in education courses is a 3.41, according to the NCES study, "Out of the Lecture Hall and into the Classroom," which compared the average grades of teaching candidates across the nation in different academic subject areas. As a point of reference, the average grade in social science courses was a 2.96, and in science 2.67. We stand by the more objective data provided by the SAT and ACT, showing that teachers who have attended teacher education programs on average have lower SAT and ACT scores.

DARLING-HAMMOND: *Walsh misunderstands the Educational Testing Service study reporting on the performance of teachers from NCATE-accredited schools, failing to acknowledge that graduates of NCATE-accredited schools of education did better on the Praxis than graduates from non-NCATE-accredited schools of education.*

RESPONSE: The report's text stands as is, with the assertion that ETS made an error in the table in question. Table 10 on page 25 of the study purports to examine whether graduates of NCATE-accredited teacher programs are better than non-NCATE accredited teacher programs, but it includes a large number of prospective teachers who never attended a teacher education program. Their inclusion skews the findings.

DARLING-HAMMOND: *Walsh's assertion about private schools not requiring certification is not as true as she leads readers to believe; the differences are exaggerated.*

RESPONSE: None of the limited data Darling-Hammond presents to refute this statement altars the fact that there are fewer certified teachers teaching in private schools than there are in any public school system. Further, the relatively low percentage of uncertified teachers employed by poor, urban districts cannot be cited as evidence of the proof of certification's value without also considering converse findings, which cannot account for why the nation's most elite private schools employ the lowest percentage of certified teachers.

DARLING-HAMMOND: *More teacher education is better, as evidenced by the fact that teachers who completed 5-year teacher training programs are more effective than teachers who attended traditional 4-year programs (page 25).*

---

[4] National Center for Education Stastistics, September 1996, *Baccalaureate and Beyond*, NCES 96899; nces.ed.gov/pubs/96899.

RESPONSE: Neither of the two studies that Darling-Hammond cites as evidence of the effectiveness of five-year programs uses student achievement as the measure of teacher effectiveness (Andrew and Schwab, 1995; Denton and Peters, 1988).

ANALYSIS OF STUDIES

Darling-Hammond has been challenged numerous times by economists for misrepresenting the findings of studies (Ballou and Pogursky, 1997, 2000; Goldhaber and Brewer, 2001). She also takes a more lax approach than most academics are willing to take when reviewing research, routinely and without qualification citing unpublished studies, doctoral dissertations, articles published in journals which have no standard of peer review, and even school board minutes.[5] In her response to the Abell report, Darling-Hammond notes that she has restricted her reference to any studies to only peer-reviewed studies. This is the first time she has ever applied that restriction to her literature reviews.

DARLING-HAMMOND: ***Wenglinsky (2000)*** *shows that students whose science teachers have had more pre-service training in science methods have higher achievement (page 7).*

RESPONSE: Wenglinsky never mentions pre-service training of teachers. He did not measure its impact on student achievement. The only teacher input that Wenglinsky measured was the effect of a teacher's major or minor in subject area on student NAEP scores. Wenglinsky expressly notes this fact on page 31: "Other inputs not included in this study, such as the preservice training of teachers or their proficiency in pedagogical knowledge as measured by standardized test scores might very well make a difference." The observation though is purely speculative.

DARLING-HAMMOND: *Walsh misunderstands some fundamental research design issues, including the difference between experimental and correlational studies. (Walsh dismissed many of the studies cited by certification advocates because they used very small sample sizes.)*

RESPONSE: In framing this argument Darling-Hammond erroneously defines an "experimental" study. She cites **Hawk, Coble and Swanson** as a quasi-experimental study, which it is not.

PODGURSKY: Hawk, Coble, and Swanson, published in Journal of Teacher Education, assess the teaching performance of 36 teachers, 18 of whom are mathematics certified and 18 of whom held certification in areas other than mathematics. They use a "paired-comparisons" method:

"An in-field teacher was paired with an out-of-field teacher who was in the same school, teaching the same mathematics course to students of the same general ability level."

The students of these teachers were pre-tested and then later re-tested in general mathematics or algebra. (The authors claim that a total of 826 students participated in the study, but test scores are reported for only 613 students) In addition, two classroom visits were made and these teachers were scored on performance assessment instrument. We are provided no data as to the number of schools or school districts. The schools were in North Carolina.

---

[5] In her response to Goldhaber and Brewer's 2000 study, Darling-Hammond cites an appendix to school board minutes from the Texas Education Agency (1993) as evidence that principals, supervisors, and colleagues tend to rate recruits from alternative programs less highly on their instructional skills (page 25).

BEST COPY AVAILABLE

The authors find that the post-test scores of the students of certified teachers were higher than the post-test scores of the non-certified teachers in both general mathematics and algebra. They also tested administered arithmetic and elementary algebra tests to the certified and non-certified teachers. The certified teachers scored higher on both tests. Finally, the certified teachers outscored the non-certified teachers on a classroom assessment instrument.

Darling-Hammond describes this study as "quasi-experimental." But it would only be experimental if Hawk et al. had control over how students were assigned to teachers or how teachers were assigned to classes. Clearly they did not. They had to take the data that were available. She seems to be assuming that the students were randomly assigned. But the authors did not say that, nor did they present any evidence that was the case.

In fact, the data in the study suggests just the opposite. The students in the study were pre-tested in general mathematics and algebra. Later in the year they were given a post-test in the same area. The data reported by the authors show statistically significant differences in the pre-test scores mathematics scores between the certified and non-certified teachers. (The authors claim that the differences were not statistically significant, but the statistics reported in the table indicate otherwise.)

Given the fact that the pre-test scores of the students differed between the two groups of teachers, the right statistical test would have been to compare the gains in achievement between the two groups. Instead, the authors compare the levels of post-test scores. If we do, in fact, compare the differences in gain scores, we find that the difference in gain scores for the certified and non-certified teachers in general mathematics was only significant at 10%, and was not statistically significant for algebra.

Another shortcoming of the study is the fact that the authors failed to examine the effect of certification over and above mathematics content knowledge. On average, the noncertified mathematics teachers scored lower on arithmetic and elementary algebra exams than did certified mathematics teachers. Thus, the non-certified mathematics teachers knew less mathematics, and they were not certified, presumably because they lacked pedagogy and/or mathematics courses. (No evidence is presented on what the non-certified teachers lacked.) The obvious multivariate test would have been to see whether there was a difference between certified and non-certified teachers after controlling for the teacher's mathematics test score. The authors did not do this.

In short, this study did not have a "reasonably well-controlled quasi-experimental design." It was not "quasi-experimental" at all, as it lacked statistical controls for student and teacher characteristics, and failed to adjust for differences in the pre-test scores of students. It is a suggestive and interesting study, but it is small and not "well-controlled."

DARLING-HAMMOND: *Walsh only reluctantly reported **Monk**'s findings of a positive effect from methodology coursework in science on student achievement. Walsh did so only after Monk corrected her and even then, she was only willing to cite Monk's findings in the appendix.*

RESPONSE: Full and fair consideration was given to Monk's findings (more consideration than it may have ultimately deserved after reviewing Podgursky's analysis). On page 7 of the on-line version of the study, Monk was cited in the section of the report devoted to firm findings about teacher quality, an important citation that Darling-Hammond did not acknowledge.

Since we released the on-line version of the study, we have reconsidered the rigor of Monk's findings. The published version of the study reflects this revision.

PODGURSKY: Monk (1994) examines the effect of teacher mathematics and science coursework and educational credentials on student achievement scores in mathematics and science. His data are from a longitudinal data file developed by the National Science Foundation (Longitudinal Study of American Youth) that tracks for two years a nationally representative 10th grade cohort of students beginning in fall 1987.

These survey data are unique in that the mathematics and science teachers were asked fairly detailed questions about the course preparation and academic background in mathematics and science. They permit Monk to estimate the effect of mathematics and science content courses and mathematics and science education courses on student achievement. Monk finds some evidence that both content and pedagogy courses matter. In her rejoinder, Professor Darling-Hammond chose to emphasize two findings in this study: a) that there is evidence of diminishing returns to mathematics content courses (more than five courses do not contribute appreciably to student achievement), and b) the estimated effect of an additional math content course is smaller than the estimated effect of an additional mathematics pedagogy course.

(Note that these results apply only to mathematics, not science. Darling-Hammond's response leaves the impression that these results hold for both.) While this a carefully done study with many interesting findings, I think that the results highlighted by Darling-Hammond must be treated with some caution.

1. Diminishing returns. These results are suggestive, but they must be viewed with caution. First, from a statistical point of view they are not strong. The t-statistics on the >5 mathematics course coefficients for the sophomore and junior year are -1.58 and -1.63, respectively. That means we can reject the null hypothesis of no effect only at 10 percent, and then only on a one-tailed test. In fact, these results are below the conventional thresholds for significance in a study based 1492 and 983 observations respectively. Moreover, even these t-values are probably overstated since there is no correction of the estimated standard errors for clustering of the observations in the 51 school sites sampled in the study. Corrections for this type of clustering are now commonplace in published econometric studies.

Darling-Hammond fails to mention the perverse results for science. Here Monk finds just the opposite result: increasing returns to physical and biological science courses. Monk discusses a variety of possible explanations for the odd result, including the possibility that physical science course-taking may be a proxy for the general intelligence of the teacher (a variable that was unavailable in these data). But this highlights the fact that other results may be affected by the lack of controls for general ability of the teacher.

Finally, Monk reports that some teachers apparently misinterpreted the question on coursework and reported credit hours rather than courses. Researchers who developed the LSAY attempted to correct this problem in later releases of the survey. However, to the extent that the problem was not fully corrected, this type of measurement error could readily produce a spurious curvilinear (diminishing returns) relationship.

To summarize, these curvilinear results are interesting. However, the result for mathematics is only weakly significant. The result for physical science is highly significant and perverse. These results have not been replicated in other studies. Hence, they must be treated with caution. To inject such tentative results into education policy debates is inappropriate.

2. Larger marginal effects for science and mathematics pedagogy as compared to content courses.

First, this result only holds for mathematics. The marginal effect of an additional mathematics course is smaller than the marginal effect of an additional math ed course at the sophomore and junior years. However, for the sciences, this effect is reversed. The effect of physical science courses is larger than the effect of science ed courses.

It should also be noted that teacher mathematics coursework was associated with higher student achievement in both science and mathematics. Mathematics courses had a consistent positive effect on the science test scores of students. The point estimates of this effect were actually bigger than the effects of science education courses in the junior year, although the difference is probably not statistically significant.

The last point highlights the fact that in a complicated empirical study like this, it is easy to pick and choose what one finds attractive and ignore other results. For example, the most powerful predictor of student science achievement at the junior level is whether the teacher was a science major. However, this same variable is statistically insignificant at the sophomore level. This type of instability in coefficients is another reason for caution.

Taken as a whole, this is an interesting and provocative study. It should stimulate further research in the area. There is some evidence in this study that in both science and mathematics teacher coursework in the discipline and in pedagogy seems to be associated with higher student achievement. However, the points emphasized by LDH on the comparative size of effects and curvilinear relationships must be viewed with caution.

DARLING-HAMMOND: *Walsh's "mysteriously" objects to the strategy used in* ***Druva and Anderson's*** *meta-analysis where they combined various teacher attributes, including education coursework, into one variable.*

RESPONSE: My objection was that the authors implied, as well as did Darling-Hammond, that a significant positive effect for education coursework was found in 47 out of a 65 studies. In fact, education coursework only had a positive effect in 3 out of 65 studies reviewed. To combine several variables, including teacher experience and college GPA, to yield a finding that is essentially misleading and which is later cited by others using even more misleading terms is a troubling practice I encountered too frequently.

PODGURSKY: *Druva and Anderson* , published in Journal of Science Teaching, is a meta-analysis of studies of the effect of teacher characteristics on science teaching. Before getting into the methodology of this study, it is important to note the weak quality of the research it purports to summarize. There are 65 studies. Of these 65 studies, 52 are dissertations, 11 are journal articles, and two are unpublished papers. Moreover the journals considered are not rigorous psychology journals but rather education journals (Journal of Research in Science Teaching and Science Education).

Many researchers have noted the difficulty of establishing causal relationships between teacher characteristics and student achievement. The teachers and students are not randomly matched to one another and socioeconomic factors are strongly associated with student achievement. Thus, it is extremely important that a study of teacher effects have good controls for student characteristics. Druva and Anderson did not require a rigorous study design or SES controls to appear in this literature review (in contrast to Hanushek, and Hedges, Greenwald, and Laine.) Druva and Anderson merely required that the study report a correlation coefficient. The authors are aware of this limitation. They consistently describe the relationships as "associations" or "correlations." Only in once sentence of the conclusion do they venture a causal statement.

Given the dubious quality of the literature being surveyed, any conclusions derived from this study must be viewed skeptically. A synthesis of research is no stronger that the literature on which it is based. Computing the central tendency of a collection of methodologically flawed studies does not add to our knowledge of the world. <u>This is a very weak study and as such should not be used for formulating education policy.</u>

With this disclaimer in mind, Darling-Hammond proceeds to misrepresent and exaggerate what is in this weak study. Only <u>three</u> -- not 47 -- studies examine the relationship between the number of education courses and "teaching effectiveness," a term that is vaguely defined by Druva and Anderson with no indication that they looked at student achievement ("The ability to produce desired change within the classroom as perceived by students and principals.")

The "47 studies" to which Darling-Hammond refers arise only when Druva and Anderson bundle number of education courses with GPA, student teaching grades, and four measures of teaching experience. When we turn to the 23 studies that measure the correlation between student achievement and this bundled "education and performance measure" that combines education coursework with GPA and experience, the average correlation is only .10. By contrast, in the 24 studies that measure the correlation between science training (as measured by science coursework) and student cognitive measures the average correlation is nearly twice as large at .19.

(By comparison, the "heterosexuality" and "masculinity" of the teacher was much more strongly associated with student achievement than either of these variables; what should be the policy implications of this finding?)

In short, this is a meta-analysis of what is arguably a very weak body of research. Nonetheless, an objective reading of the results of this study finds stronger support for the importance of science content knowledge as compared to pedagogical training. Darling-Hammond can only arrive at her conclusion by misrepresenting a bundled "education and performance" variable as primarily a measure of education coursework. A careful reading of the study shows that it is not.

DARLING-HAMMOND: *Walsh should have cited **Begle's** studies as evidence of the strong effect of mathematics methodology coursework. Walsh selectively cites evidence, reporting that Begle's study had found that mathematic courses make a difference, but she dismisses his findings on methodology coursework.*

RESPONSE: This argument might have some merit if Abell's report had cited Begle as evidence of a strong effect from mathematics coursework, but it did not. Begle's study was too problematic to use for any purpose. In an email correspondence with Begle's editor, James Wilson (Begle is deceased), he noted that Begle's findings should be viewed with great skepticism because the design of the study is not defendable. Wilson stated:

> "The underlying problem was that there were so many variables and so much data but not the talent, resources, or technology to do the extent of analyses, interpretation, and reporting that the mathematics, educational psychologists and methodologists anticipated. Instead we got a lot of data summaries and first level analyses."[6]

DARLING-HAMMOND: *Walsh interpreted the NTE measure in* ***Strauss and Sawyer*** *as a measure of verbal ability, to suit the purposes of Abell's argument, and does not acknowledge that it is a test of professional knowledge.*

RESPONSE: At no point did Abell contend that the NTE was not in part a test of professional knowledge. The report does note that there is no research supporting a connection between teachers' scores on the NTE and their certification status. Teachers who have had **no** education coursework can and do get higher scores on the NTE than teachers who have had education coursework, which does not speak well for the rigor of education coursework. In Maryland, there is strong evidence of this finding. Teachers who enter teaching through Teach For America and the Resident Teacher Certificate have higher Praxis scores (the successor of the NTE) than the average teacher score in the state.

DARLING-HAMMOND: *Walsh dismisses the evidence found in* ***Ferguson and Womack*** *(1993).*

RESPONSE: Ferguson and Womack (1993) is a study of 266 graduates of the teacher education program at Arkansas Tech University. This study adds little value to the debate over certification, as it does not study any uncertified teachers.

PODGURSKY: *Ferguson and Womack* is a study of 266 secondary student teachers from Arkansas Tech University. These student teachers were evaluated using a variety of instruments during classroom visits. There we no measures of student achievement. The assessment evaluations were then related to course grades in six pedagogy courses (e.g., Introduction to Teaching, Human Development, Methods of Instruction), overall GPA and NTE specialty test scores. The authors find that course grades in the six pedagogy courses as a group explained more of the variation in assessment scores than did overall GPE or NTE specialty test scores. They conclude:

"... the findings of this study indicate that coursework in teacher education makes a positive difference in teaching performance and that education coursework is a more powerful predictor of teaching effectiveness that measures of content expertise (GPA in major and NTE specialty courses). This strongly suggests that it would be counterproductive to increase content course exposure at the expense of coursework in pedagogy." (p. 61)

---

[6] Email correspondence with James Wilson.

So far as I can determine, this conclusion is based entirely on a flawed statistical methodology. The authors base this conclusion on the incremental contribution of GPA and NTE to the overall R-square of the statistical model. However, it is well known that the incremental contribution of an explanatory variable to R-square depends on the order in which variables are entered in the model. These researchers control for the grades of the six pedagogy courses, the grade in student teaching and then enter GPA and NTE scores into the model. Not surprisingly, the incremental contribution of GPA and NTE to R-square is small (but highly significant statistically). But now reverse the experiment. Control first for GPA, NTE and student teaching grade and then enter the grades of the six classes. It will certainly be the case that the contribution to R-square of these variables will be small as well, and much smaller than the contribution of GPA and NTE.

In fact, this study shows that GPA and NTE scores correlate with the performance evaluations of student teachers. NTE scores matter even after controlling for a variety of course grades and GPA. However, the researchers' conclusions about which variables are more powerful predictors of teacher performance evaluations are incorrect.

DARLING-HAMMOND: *Walsh did not read* ***Guyton and Farokhi*** *carefully (page 36) for the ample evidence it provides of the value of certification.*

RESPONSE: The study provides little evidence no matter how good or bad one considers its methodology to be. Guyton and Farokhi is a study of approximately two hundred graduates from the school of education at Georgia State University in the early 1980's. It does not provide any insight as to the possible gains that may be gained from hiring liberal arts majors as teachers. It contains no data on student achievement. Basically, the conclusion of the study is that if you are going to hire a teacher who was an education major, you should hire one who had a high GPA.

PODGURSKY: *Guyton and Farokhi,* published in Journal of Teacher Education, the researchers track several hundred education graduates of Georgia State University who took teaching jobs in Georgia. The measure of teacher performance in this study is an evaluation of new teachers administered by the Georgia education department. It involved classroom visits (presumably announced, we are not told) and an analysis of teacher portfolios. There are 14 competency scores in the performance assessment. There are no measures of student achievement.

The independent variables are tests of teachers' basic knowledge (Regents Essay and Reading exams), sophomore GPA, certification exam score, and upper level GPA (ULGPA). The authors find ULGPA more strongly correlated with the competency scores than any of these other variables. They interpret this to mean that what students learn in pedagogy courses is the most important predictor of performance, as measured by the performance assessment.

Even if we accept the result at face value, it still does not tell us much. The performance assessment measures things that ed school students are supposed to learn to do in ed school. It is not surprising that students with higher ed school GPA's would get better scores on such an assessment.

The authors report simple correlations. One problem is that there are very different sample sizes depending on which two variables they are correlating. For example in correlating ULGPA and the performance assessment they have n=269, while the sample for the regents reading exam

and performance assessment is only n=151. Even if these were random exclusions, that in itself can explain part of the lower significance level for the latter. But these are not random exclusions. And the authors do not explore this issue. At a minimum they should have reported a correlation matrix for exactly the same set of observations (n=151).

Another problem that the authors mention is attenuation (p.39). The basic problem is that if there is a minimum regents exam score or minimum sophomore GPA for admission to an ed program, or a minimum certification score for licensing, then you are eliminating some of the variation in the independent variable. It can be shown that this will weaken the correlation in the dependent variable and the attenuated variable. This is particularly a problem for one of the general knowledge scores. The essay score ranges from 1 to 4 and you need a 2 to pass. There apparently is very little variation in this variable. The mean is 2.03 but the standard deviation is only .18, that is, there is very little variation. If everyone basically has the same score, it really can't predict anything. However, if the ed students who got a one on the writing exam had been allowed to be teachers we might have found a large correlation. But they were weeded out of the pool at an earlier stage.

In sum, the primary finding of this study is that classroom assessments of a sample of new teachers who graduated from the Georgia Tech education school were correlated with their ULGPA. This result is of some interest, but at best it tells us that if one is going to hire ed school graduate, it's better to hire one with a higher ULGPA. It does not, for example, tell us whether these education majors were better than liberal arts majors. Nor do the researchers provide us with any data on student achievement gains of these teachers' students.

DARLING-HAMMOND: *Alternative teachers in the* ***Miller, McKenna and McKenna*** *study had to take a lot of education coursework, "15 to 25 credit hours" a requirement that is as extensive as what Maryland requires for traditionally certified teachers (page 11).*

RESPONSE: Darling-Hammond misreads Miller et al.'s study. The alternative certification teachers had to take 9 to 15 semester credits, in addition to intensive mentoring, in-service training and supervision. We have no way of knowing which component of this program was responsible for the teachers' success. I did not state that its sample size was insufficient, as Darling-Hammond claims; it was a matched pair study so its small numbers are justified. None of the other studies that I suggest may be too small used this gold standard of research.

DARLING-HAMMOND: *Abell should not have dismissed* ***Schalock*** *because of it citing old research, when the report cites even older studies to provide evidence to the contrary.*

RESPONSE: When I contacted Schalock, he wanted to know why I would bother wanting to look at his "old, old!" review of the literature, most of which looked at IQ research done in the 1940s. Schalock wrote his paper in 1979; he was not presenting original research but reviewing literature *from the 1940s*. The studies he discusses are certainly much older than any studies that I cited; the dates on the verbal ability findings that I cite range from 1971 to 1996.

ERRORS

Given that over 200 studies were reviewed for this report, Abell is pleased with the few number of errors that our report's most vocal critic identified. Where Darling-Hammond found any errors, either in the on-line report, or in the August 2000 *draft* of the paper sent to Darling-Hammond, they were corrected in the hard-copy version. (Abell gave Darling-Hammond a draft of the report a few months ahead of its publication, so that she could make comments and corrections.)

1. Darling-Hammond finds a misprint in the draft appendix that attributes the findings on the importance of teacher's verbal ability to Ferguson and Womack (1996). The correct citation should have been Ferguson and *Ladd*.

2. I cite two studies which were not peer reviewed to support findings against certification: these studies have been deleted from the published version so that the report is consistent and does not appear to convey a double standard. I also cited two studies that were not peer reviewed as evidence that SAT scores are lower for teachers (a fact that Darling-Hammond does not challenge) and has deleted these two citations as well. However, Darling-Hammond's assertion that I did so 15 times is unfounded.

3. I cite five studies (on pages 7 and 8) which had not used student achievement as the measure of teacher effectiveness, a standard that I imposed for the research that the report would consider legitimate. Darling-Hammond is correct; these studies should not have been included and have been deleted in the printed version.

4. I attribute a chart at the beginning of Section 3 to a study done by Greenwald, Hedges, and Laine which should have been attributed to Ferguson.

5. Darling-Hammond (1999) did consider class size in her NAEP analysis, though I noted mistakenly said that she did not. The printed version reflects this correction, noting that Darling-Hammond did not consider other important factors, such as race, in the analysis.

DARLING-HAMMOND: *Walsh misrepresents conversations and analyses that she had with authors.*

RESPONSE: Darling-Hammond paints an impression that I misled some of the researchers whom I interviewed, misrepresenting their comments or causing them to say something that they would not have said had they known the intent of Abell's report (Mark Fetler, David Monk, and Larry Hedges). There is nothing to be gained by reviewing the details of these conversations, though Darling-Hammond retells them inaccurately. The report's analyses of their three studies are accurate; the content of their studies is the issue, not the nature of any comments the researchers may have made to me or to Darling-Hammond. I invite the reader to verify the accuracy of our analyses by reading their studies.

DARLING-HAMMOND: *Walsh dismisses old studies simply because they are old.*

RESPONSE: I did not eliminate studies just because they were old. I do present the problems that occur when depending exclusively on old studies as evidence, but I state clearly (page 21) that just because research is old is not reason alone to dismiss it. Further, I was far more flexible than others have been on accepting older research. In a literature review that was published last year that was sympathetic to certification, the researchers *automatically* dismissed any study that was more than twenty years old (Wilson, Floden, Ferrini-Mundy, 2001).

Darling-Hammond similarly misrepresents my mentioning studies where aggregation bias may be a problem. I did reject studies where aggregation bias was *clearly* a problem (such as Darling-Hammond's NAEP state-wide analysis); for many studies, I merely point out where aggregation bias *might* be a problem, often repeating the researcher's own concerns about the study.

DARLING-HAMMOND: *"It is ridiculous to argue that knowledge of teaching and learning and the opportunity to learn to teach under the close supervision of a master teacher through student teaching and other guided experiences do not matter at all" (page 36).*

RESPONSE: The Abell report never argues that these experiences are incapable of having value. Probably every new teacher can benefit from the guidance of a master teacher (provided the master teacher is any good); many prospective teachers have found it beneficial to serve as a student teacher as well. Abell's point is the following: the value that may be gained by a pre-service experience such as student teaching does not justify barring prospective teachers from the profession, simply because they have lacked such pre-service experiences.

DARLING-HAMMOND: *Abell is recommending that states collect verbal ability scores only on prospective teachers who have gone to schools of education.*

RESPONSE: This is not accurate. Abell's recommendation applies to all teachers no matter what their background.

DARLING-HAMMOND: *Principals are not in the best position to have control over teacher hiring because a principal is not in the position to control teacher supply/demand issues, and other consequences of deregulation.*

RESPONSE: Darling-Hammond mischaracterizes the report's recommendation. Abell's recommendation was that the state turn over the decision making to school districts, which in turn should be encouraged to give school principals the flexibility to hire as they see fit (as private school principals do), provided adequate accountability measures for principals are in place. School districts would still address all of the supply/demand issues raised by Darling-Hammond, such as noncompetitive teacher salary levels, much as they do now.

The concerns that Darling-Hammond raise here have much more to do with job protection, teacher pay and assorted union issues than how to improve student achievement. She seems to be asserting that if we let schools hire uncertified teachers this will reduce pressure to increase teacher pay. Even it this were true (and there is no indication that it would be true), schools operate to educate children. If the issue is whether higher teacher pay brought about by certification entry barriers is a cost-efficient way to improve student achievement, then the first step in that analysis should be an assessment as to whether teacher certification is associated with higher levels of student performance. Our analysis indicates that it is not.

Darling-Hammond makes a somewhat confusing statement that "In addition, eliminating certification requirements would eliminate evidence about disparities in student's opportunities to learn, for if there were no minimum standards, there will be no evidence of differences in the extent to which they have been achieved by teachers working with different groups of students." Given that this statement is couched in legal terms, she *seems* to be arguing that if districts are allowed to hire uncertified teachers, it will be harder to win school finance lawsuits proving that districts hiring more uncertified teachers are actually harming students. If the goal of state education policy makers is to make it easier for themselves to be sued, Darling-Hammond has a point here.

DARLING-HAMMOND: *The Abell report does not distinguish between teachers who are not certified at all and teachers who are alternatively certified.*

RESPONSE: Darling-Hammond has some of the same problems I did with wrestling with various research findings that looked at different certification policies, including alterative certification policies, found in 50 different states. For example, she labels teachers uncertified in Goldhaber and Brewer's 2000 study who were in fact certified, but teaching out of their field. But Abell's findings are not altered because of this problem. I did not uncover a single study, including the 19 that Darling-Hammond examined in her response, providing persuasive evidence that certified teachers produce higher student achievement gains. If state regulations are to have a purpose, if states want to use regulatory policy to ensure high standards, then the current process for doing so is demonstrably ineffective. Absent what should be requisite evidence, states need to look for other means to ensure greater teacher quality.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Teacher Certification Reconsidered: Stumbling for Quality (0579)

Author(s): Kate Walsh (Kate Walsh)

Corporate Source: The Abell Foundation (The Abell Foundation)

Publication Date: December 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☐

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Anne LaFarge Coleman	Printed Name/Position/Title: ANNE LAF. COLEMAN / VICE PRESIDENT
Organization/Address: The Abell Foundation	Telephone: 410-547-1300 / FAX: 410-539-6579
	E-Mail Address: / Date: August 22, 2003

Suite 2300
111 South Calvert St.
Baltimore, Md. 21202

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Teaching
and Teacher Education
1307 New York Ave., NW
Suite 300
Washington, DC 20005-4701

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)